Defeating Certain Types of AI-Based Suspicious Activity Detection Using Human-Emulative AI to Help Espionage-Related Cyber Traffic Pass "Smell Test"

3 January 2024
Simon Edwards
Research Acceleration Initiative

Introduction

A great deal of focus has been placed upon heuristic detection of suspicious patterns of behavior, however, a new class of defensive algorithm; based instead upon recognition of the absence of normative cyber activity coming from nodes; has been overlooked. Entities employing this type of defense treat traffic from nodes without an ongoing history of Internet usage consistent with an authorized user as unauthorized or, at minimum, suspicious.

Abstract

Overcoming this type of defense requires an extremely intricate emulation (AI-based) of normative patterns of Internet usage by a node and the establishment of an ongoing pattern of activity prior to an intrusion attempt (Blue-on-Red.) This establishment of a normative pattern of activity prior to making even the most preliminary intrusion attempt is now an absolute requirement in the cyber domain, necessitating the dedication of non-trivial resources to the defeat of this innovative security precaution pioneered by PLA-3.

This new paradigm requires that fully autonomous AI programs run for a period of weeks or months at the ISP level (in China, for example) where they masquerade as authentic Internet subscribers. A hybridization of observed patterns of traffic of other users would then be used by this bot in order to order to establish the node as trustworthy. One challenge of making these patterns of activity realistic is that entire families must be simulated (a typical node would include multiple devices accessing the Internet from a home router and with unique search histories appropriate to the age and gender of the members of the family.)

If, for example, one of these simulated nodes had a search history inconsistent with a middle-aged male PLA-3 member, it would not pass muster with the established AI-based suspicious activity detection algorithms currently in use.

Conclusion

In order to estimate the efficacy of any given algorithm/AI along these lines, Blue-versus-Red gaming would be required in order to determine whether American versions of this defensive technology are able to differentiate simulated Internet users from authentic users. Should this endeavor fail, activities internal to entities such as PLA-3 risk becoming a black box from an outsider's perspective. The longer this remains the case, the more difficult it will

be to gain access to new, authentic information concerning PLA-3 cyber defenses.